Examples of linear regression

Manel Martínez-Ramón

ECE, UNM

October, 2018

In this example we apply a linear GP regression over the linear model

$$y_n = 0.5x_n + 0.5 + w_n$$

where x_n is a uniform random variable between 0 and 1 and w_n is an iid Gaussian noise with variance σ^2.

Training:

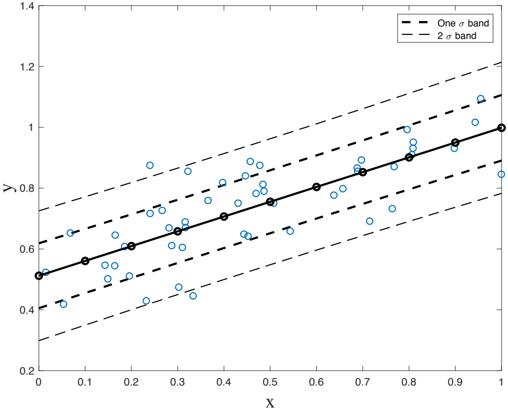
- 50 samples x_n, y_n taken from the linear model above.
- $\sigma = 0.05$
- The linear GP model is initialized with $\sigma_n = 0.9$.
- σ_n is further validated to maximize the training data log likelihood (to be presented in next lessons).

Test:

- 10 samples spaced between 0 and 1.

Example: One dimension regression

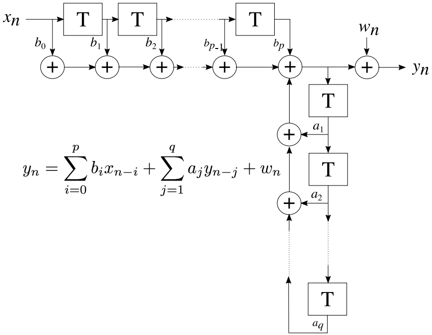
Results



Number of samples in the σ band: 16 (68%). Theoretical value: 68.3%
Number of samples in the 2σ band: 2 (96%). Theoretical value: 95.5%
Statistics taken from training samples.

Example: linear prediction

In this example, a prediction of the output sample $y[n]$ of a linear system from past samples $\mathbf{y}[n-1]$ (defined in the next slide). The system is depicted in the diagram.



$$y_n = \sum_{i=0}^{p} b_i x_{n-i} + \sum_{j=1}^{q} a_j y_{n-j} + w_n$$

The output is a linear combination of input signals x_n to x_{n-p} and past values of the output. We will predict output y_n from past samples y_{n-j}, $1 \leq j \leq q$

Prediction structure

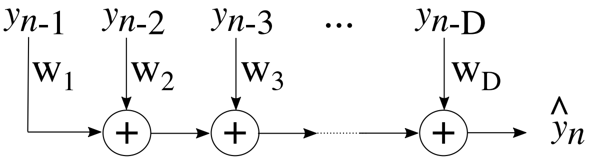
The prediction $\hat{y}_n = f(\mathbf{y}_{n-1})$ of output sample y_n is performed using samples

$$\mathbf{y}_{n-1} = [y_{n-1}, \; \cdots, y_{n-D}]^\top$$

so, the prediction function can be written as

$$\hat{y}_n = f(\mathbf{y}_{n-1}) = \mathbf{w}^\top \mathbf{y}_{n-1}$$

which can be represented as in the figure.



Note that we ignore the bias term because we assume that the output has zero mean.

Experimental setup

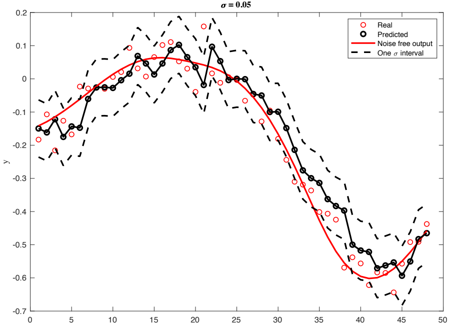
A set of N samples x_n is drawn from a Gaussian distribution with zero mean and unit variance. This data is processed with the system of slide 4. The values of the coefficients are

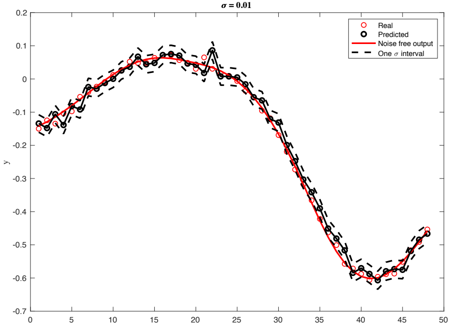
$$\mathbf{b} = [0.0312,\ 0.1250,\ 0.1874,\ 0.1250,\ 0.0312] \times 10^{-3}$$
$$\mathbf{a} = [1.0000,\ -3.5897,\ 4.8513,\ -2.9241,\ 0.6630]$$

The length of the predictor vector $\mathbf{y}_{n-1}$ is D=3. A sequence of $N = 100$ data is generated for training, and then a new sequence is using for test purposes. The noise added after the filter of slide 4 is Gaussian of zero mean and variance σ_n^2 taking various values.

The filter values have been generated with Matlab function [b,a]=butter(4,0.05), which synthesizes the coefficients of a digital lowpass Butterworth filter of order 4. Thus, $p = q = 5$.

Results





The two figures represent a the prediction of a test sequence with two levels of noise. In both cases, the counts of errors inside the one σ and 2 σ intervals are in high agreement with the theory.